Nikita Bhutani

Research Scientist
Megagon Labs, MountainView, CA

✉ nikita@megagon.ai
🏠 nikibhutani.github.io
in nikita-bhutani
G google scholar

RESEARCH INTERESTS

My research interests lie at the intersection of Data Management, Natural Language Processing and Machine learning, with special focus on evolving and emerging areas such as: knowledge-base construction, natural language understanding and question answering systems.

EDUCATION

Ph.D. in Computer Science and Engineering, University of Michigan, Ann Arbor, MI   June 2019
  Advisor: Prof. H V Jagadish
  Committee: M. Cafarella, R. Mihalcea, W. Lasecki, Yunyao Li, Qiaozhu Mei
  Thesis: Answering Complex Questions with Heterogeneous Knowledge Sources derived from Text

M.S.E. in Computer Science and Engineering, University of Michigan, Ann Arbor, MI   May 2016
  Advisor: Prof. H V Jagadish
  GPA: 4.0/4.0

B.Tech. in Textile Technology, Indian Institute of Technology, Delhi, India   May 2010
  Advisor: Prof. Manjeet Jassal
  GPA: 8.73/10 | *Rank*: 2

WORK EXPERIENCE

Megagon Labs, Mountain View, CA
  Research Scientist   August 2019 - Present
  Natural language processing, machine learning, question answering

University of Michigan, Ann Arbor, MI
  Graduate Student Research Assistant with Prof. H V Jagadish   August 2014 - June 2019
  Open information extraction, knowledge-based question answering, natural language interfaces

Megagon Labs, Mountain View, CA
  Research Intern with Wang-Chiew Tan   June 2018 - August 2018
  Open information extraction from conversational question-answer pairs

IBM Research, Almaden, CA
  Research Intern with Yunyao Li   June 2017 - August 2017
  On-demand curation of text and integration with structured KB

  Research Intern with Yunyao Li   June 2016 - August 2016
  Learning structured representations of named entities for entity resolution

Ubiquiti Consultants Pvt. Ltd., Delhi, India
Software Engineer/ Data Scientist   July 2010 - June 2014
Developed UX/UI of software suite for analytics, diagnostics and search of automative data. Curated ontologies for extracting information from semi-structured and unstructured automative data.

Technical University of Liberec, Czech Republic
Summer Intern with David Lukas   May 2008 - July 2008
Electrospinning from free liquid surfaces

HONORS AND AWARDS

- Nominated by UM-CSE for Rackham Barbour Scholarship, 2018
- Rackham Conf. Travel Grant, 2018, 2017, 2016
- IBM PhD Fellowship, 2017
- GHC Travel Scholarship, 2016
- UMich PhD Fellowship, 2014
- Merit Award (5 semesters), IIT Delhi
- Best B.Tech. Thesis, IIT Delhi

CONFERENCE/WORKSHOP PAPERS

1. SubjQA: A Dataset for Subjectivity and Review Comprehension
   Nikita Bhutani, Johannes Bjerva, Behzad Golshan, Wang-Chiew Tan, Isabelle Augenstein. *The 2021 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2020)

2. SAMPO: Unsupervised Knowledge Base Construction for Opinions and Implications
   Nikita Bhutani, Aaron Traylor, Chen Chen, Xiaolan Wang, Behzad Golshan, Wang-Chiew Tan. *Conference on Automated Knowledge Base Construction 2020* (AKBC 2020)

3. Answering Complex Questions by Combining Information from Curated and Extracted Knowledge Bases
   Nikita Bhutani, Xinyi Zheng, H V Jagadish. *Proceedings of the First Workshop on Natural Language Interfaces at ACL 2020* (NLI-ACL 2020)

4. Answering Complex Queries with Heterogeneous Structured Knowledge Sources extracted from Text
   Nikita Bhutani. *2019 Workshop on Conversational Access to Data (CAST) at SIGMOD 2019* (CAST-SIGMOD 2019)

5. Online Schemaless Querying of Heterogeneous Open Knowledge Bases
   Nikita Bhutani, H V Jagadish. *28th ACM International Conference on Information and Knowledge Management* (CIKM 2019)

6. Learning to Answer Complex Questions over Knowledge Bases with Query Composition
   Nikita Bhutani*, Xinyi Zheng*, H V Jagadish. *28th ACM International Conference on Information and Knowledge Management* (CIKM 2019)

7. Open Information Extraction from Question-Answer Pairs
   Nikita Bhutani, Yoshihiko Suhara, Wang-Chiew Tan, Alon Halve, H V Jagadish. *17th Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL-HLT 2019)

8. Exploiting Structure in Representation of Named Entities with Active Learning
   Nikita Bhutani, Yunyao Li, H V Jagadish, Kun Qian, Mauricio A. Hernandez, Mitesh Vasa. *27th International Conference on Computational Linguistics* (COLING 2018)

9. LUSTRE: An Interactive System for Entity Structuring and Variant Generation
   Kun Qian, **Nikita Bhutani**, Yunyao Li, H V Jagadish, Mauricio A. Hernandez. *34th IEEE International Conference on Data Engineering* (ICDE Demo 2018)

10. Nested Propositions in Open Information Extraction
    Nikita Bhutani, H V Jagadish, Dragomir Radev. *The 2016 Conference on Empirical Methods on Natural Language Processing* (EMNLP 2016)

11. Electrohydrodynamics of free liquid surface in a circular cleft: An application to electrospinning
    Nikita Bhutani, David Lukas. *Fiber Society Technical Conference, 2008*

PATENTS

- Entity Structured Representation and Variant Generation
  Nikita Bhutani, Yunyao Li, Mauricio A. Hernández, Kun Qian, Min Li. *U.S. Patent 10585986*
- Resolving Queries using Structured and Unstructured Data
  Nikita Bhutani, Kun Qian, Yunyao Li. *U.S. Patent Pending*

ACADEMIC SERVICE

- PC Chair: SoCC Poster 2019
- PC Member: VLDB2022, CIKM Poster/Demo 2020, CIKM Demo 2019,
- Reviewer for: IEEE Access 2020, SIGMOD 2020, AAAI 2019, IEEE BigData 2019
- External Reviewer for: TKDE 2019, VLDB 2019, VLDB 2018
- Student Mentor for: Xinyi Zheng, Aditi Sharma

SELECTED PROJECTS

- Question Answering over reviews
  Developing a robust knowledge-enhanced question answering system that answers user questions from reviews using limited training data
- Mining commonsense knowledge from reviews
  Designing pre-training tasks to teach models entity-specific knowledge and domain-specific commonsense to improve review understanding
- Salient Fact Extraction
  Developing an extractor that finds characteristic facts about specific entity from reviews
- Unsupervised Knowledge Base Construction for Opinions and Implications
  Developed a matrix-factorization method to mine opinions and their implications from review corpus
- Subjectivity in Review Comprehension
  Developed a QA dataset targeted at understanding subjective text in reviews
- Hybrid KB-QA over open and curated knowledge bases
  Developed a KB-QA system that combines both automatically extracted and carefully curated (but incomplete) information to answer complex questions

- Online schemaless querying of heterogeneous open knowledge bases
  Developed a querying method for open KBs that is agnostic about query specification and finds answers from facts having diverse representations

- Open Information Extraction from Question-Answer Pairs
  Developed a multi-encoder, constrained-decoder framework that extracts tuples from multiple sentences in a conversational question-answer pair

- Canonicalization of open knowledge bases (in collaboration with *IBM Cognitive Horizons Network*)
  Clustering entity and relation phrases to canonicalize redundant and ambiguous facts in open KBs

- Nested propositions in open information extraction
  Developed an open-domain extractor that uses bootstrapping to extract multiple complex assertions as nest-tuples from textual data with no pre-specified relations or training data

- Open information extraction from biomedical data
  Developed an open-domain extractor that extracts nest-tuples from biomedical literature

- Template-based NLI for relational databases
  Developed an NLIDB system that models natural language queries as a set of weighted SQL templates describing the likely query logics and their likelihood to be queried

- Representing news articles as RDF triples
  Developed a rule-based system to extract nominal, temporal, spatial, event-based relations between entities in news articles

- Optimizing loop unroll factors using machine learning
  Developed a supervised learning approach to identify profitable loop candidates and optimal unroll factors

- Melt electrospinning of nano-fibres (*B.Tech. Thesis*)
  Designed and developed the first in-house prototype for melt electrospinning of nano-fibres, as part of the largest and highest funded research group at IIT Delhi.

OTHERS
- NLP contributor: Climate Mind