

Luca Soldaini

luca@soldaini.net

EDUCATION

Georgetown University

Doctor of Philosophy (Ph.D.) in Computer Science

Washington, DC, USA

August 2013 – April 2018

- **Research interests:** information retrieval, machine learning, and natural language processing.
- **Dissertation:** “*The Knowledge and Language Gap in Medical Information Seeking.*”
- **Adviser:** Nazli Goharian.
- **Committee:** Der-Chen Chang, Ophir Frieder, Nazli Goharian, Elad Yom-Tov, Wenchao Zhou.

Georgetown University

Master of Science (M.S.) in Computer Science; GPA: 4/4

Washington, DC, USA

August 2013 – May 2015

Università degli Studi di Firenze

Bachelor of Engineering (B.Eng.) in Computer Engineering; GPA: 27.7/30

Florence, Italy

September 2009 – April 2013

- **Thesis:** “*Particle Swarm Algorithm for Sphere Packing Problems.*”
- **Adviser:** Fabio Schoen.

EMPLOYMENT

Amazon – Alexa Machine Learning Team

Applied Scientist

Cambridge, MA, USA

June 2018 – Present

- Working on improving components of the natural language understanding pipeline for Amazon Alexa devices.

Applied Research Intern

June 2017 – August 2017

- Studied the problem of obtaining multilingual, domain specific word embeddings that can be used to accelerate model training in new languages.
- Explored methods for bootstrapping named entity recognition to new languages when no domain-specific training data is available.

Microsoft Research – Advanced Technology Labs Israel

Research Intern

Herzliya, Israel

September 2015 – December 2015

- Studied the problem of identifying small cohorts of search engine users who might be affected by the same disease (a publication based on this work has been accepted at WWW 2017).

MedStar Institute for Innovation (MI2)

Intern

Washington, DC, USA

May 2015 – August 2015

- Developed a pipeline to extract human factors concepts from patient safety events generated by care providers.
- Helped creating a system to evaluate the quality of reports produced by radiology residents (a publication based on this work has been accepted at the DMMH workshop at SDM 2016).

PEER-REVIEWED PUBLICATIONS

- Sean MacAvaney, Andrew Yates, Arman Cohan, [Luca Soldaini](#), Kai Hui, Nazli Goharian, and Ophir Frieder. “*Overcoming Low-Utility Facets for Complex Answer Retrieval.*” KG4IR Workshop, ACM conference on Research and Development in Information Retrieval (SIGIR). 2018.
- Arman Cohan, Bart Desmet, Andrew Yates, [Luca Soldaini](#), Sean MacAvaney, and Nazli Goharian. “*SMHD: a Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions.*” Conference on Computational Linguistics (COLING). 2018.
- [Luca Soldaini](#), Timothy Walsh, Arman Cohan, Julien Han, and Nazli Goharian. “*Helping or Hurting? Predicting Changes in Users’ Risk of Self-Harm Through Online Community Interactions.*” CLPsych Workshop, Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). 2018.

- Sean MacAvaney, Bart Desmet, Arman Cohan, [Luca Soldaini](#), Andrew Yates, Ayah Zirikly, and Nazli Goharian. “TempMH: Temporal Annotation of Self-Reported Mental Health Diagnoses.” CLPsych Workshop, Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). 2018.
- Sean MacAvaney, Andrew Yates, Arman Cohan, [Luca Soldaini](#), Kai Hui, Nazli Goharian, and Ophir Frieder. “Characterizing Question Facets for Complex Answer Retrieval.” ACM conference on Research and Development in Information Retrieval (SIGIR). 2018.
- Sean MacAvaney, [Luca Soldaini](#), Arman Cohan, and Nazli Goharian. “Tree-LSTMs for Scientific Relation Classification.” SemEval Workshop, Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). 2018.
- [Luca Soldaini](#), Andrew Yates, and Nazli Goharian. “Denoising Clinical Notes for Medical Literature Retrieval with Convolutional Neural Model.” Conference on Information and Knowledge Management (CIKM). 2017.
- [Luca Soldaini](#), Andrew Yates, and Nazli Goharian. “Learning to Reformulate Long Queries for Clinical Decision Support.” Journal of the Association for Information Science and Technology (JASIST), Special Issue on Biomedical Information Retrieval. 2017.
- [Luca Soldaini](#) and Elad Yom-Tov. “Inferring Individual Attributes from Search Engine Queries and Auxiliary Information.” Wide World Web conference (WWW). 2017.
- [Luca Soldaini](#) and Nazli Goharian. “Learning to Rank for Consumer Health Search: a Semantic Approach.” European Conference on Information Retrieval (ECIR). 2017.
- [Luca Soldaini](#) and Nazli Goharian. “QuickUMLS: a Fast, Unsupervised Approach for Medical Concept Extraction.” MedIR workshop, ACM conference on Research and Development in Information Retrieval (SIGIR). 2016.
- Arman Cohan, [Luca Soldaini](#), and Nazli Goharian. “Identifying Significance of Discrepancies in Radiology Reports.” Workshop on Data Mining for Medicine and Healthcare (DMMH), SIAM International Conference on Data Mining (SDM). 2016.
- [Luca Soldaini](#), Andrew Yates, Elad Yom-Tov, Ophir Frieder, and Nazli Goharian. “Enhancing Web Search in the Medical Domain via Query Clarification.” Information Retrieval Journal, 2016.
- Arman Cohan, [Luca Soldaini](#), and Nazli Goharian. “Matching Citation Text and Cited Spans in Biomedical Literature: a Search-Oriented Approach.” Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). 2015.
- [Luca Soldaini](#), Arman Cohan, Andrew Yates, Nazli Goharian, and Ophir Frieder. “Retrieving Medical Literature for Clinical Decision Support.” European Conference on Information Retrieval (ECIR). 2015.
- Arman Cohan, [Luca Soldaini](#), Andrew Yates, Nazli Goharian, and Ophir Frieder. “On Clinical Decision Support.” ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB). 2014.

TECHNICAL REPORTS

- [Luca Soldaini](#), Will Edman, Nazli Goharian. “Team GU-IRLAB at CLEF eHealth 2016: Task 3.” Conference and Labs of the Evaluation Forum (CLEF). 2016. (best submission out of 10 participants)
- [Luca Soldaini](#), Arman Cohan, Andrew Yates, Nazli Goharian, and Ophir Frieder. “Query Reformulation for Clinical Decision Support Search.” Text REtrieval Conference (TREC). 2014.
- Arman Cohan, [Luca Soldaini](#), Saket S.R. Mengle, and Nazli Goharian. “Towards Citation-Based Summarization of Biomedical Literature.” Text Analysis Conference (TAC). 2014.

TEACHING EXPERIENCE

Georgetown University

Co-instructor

- **Health search and mining** (graduate courses): Spring 2017.
- **Text mining** (graduate course): Fall 2017.

Washington, DC, USA

January 2017 – December 2017

Georgetown University

Teaching Assistant

- **Information retrieval** (undergraduate & graduate course): Fall 2013, 2014, 2016; Spring 2018.
- **Information systems** (undergraduate course): Spring 2014.
- **Data mining** (undergraduate course): Spring 2014, 2015, 2016, 2017; Fall 2017.
- **Introduction to databases** (undergraduate course): Spring 2015, Spring 2018.

Washington, DC, USA

August 2013 – April 2018

PATENTS

- Ratwani, Raj, Allan Fong, Ross Filice, Arman Cohan, Luca Soldaini, Nazli Goharian, and Ophir Frieder. "Systems and methods for targeted radiology resident training." U.S. Patent Application 15/410,850, filed January 20, 2017.

PROFESSIONAL ACTIVITIES

- **Student research supervisor:**
 - **Will Edman** (undergraduate student) "Search systems for consumer-oriented medical information retrieval." NSF REU project. Advisor: Nazli Goharian. 2016.
 - **Julien Han, Timothy Walsh** (master students) "Predicting changes in users' risk of self-harm through online community interactions." Research project. Advisor: Nazli Goharian. 2018.
 - **Ziling Fan** (master student) "Relation extraction for protein-protein interactions affected by mutation." Master thesis. Advisor: Nazli Goharian. 2018.
- **Reviewer:** Journal of the American Medical Informatics Association (JAMIA). 2017-current.
- **Program committee member:** ACL 2018 (main track), SIGIR 2018 (full and short paper tracks), CoNLL 2017 (main track), CIKM 2017, CIKM 2018 (short papers track), WWW 2017, WWW 2018 (computational health track.)
- **Subreviewer:** AAAI 2017, AAAI 2018 (main track.)

AWARDS

- **SIGIR student travel grant.** CIKM 2017.
- **Graduate School conference travel award.** Georgetown University. Academic years 2016-2017 & 2017-2018.
- **Student travel grant.** MedIR workshop. SIGIR 2016.
- **Second place at best poster award** (2 out of 40). "On Clinical Decision Support". Informatics Symposium at Georgetown University 2014.