Luca Soldaini

luca@soldaini.net

Education

Georgetown University Washington, DC, USA

Doctor in Philosophy (Ph.D.) in Computer Science Aug 2013 - Current

- Research interests: information retrieval, machine learning, and natural language processing.
- Dissertation topics: methods for semantic query reformulation, with applications in consumer health search and expert medical information retrieval.
- Adviser: Dr. Nazli Goharian.

Master of Science (M.S.) in Computer Science Aug 2013 - May 2015

- GPA: 4/4

Università degli Studi di Firenze Florence, Italy

Bachelor of Engineering (B.Eng.) in Computer Engineering Sep 2009 - Apr 2013

- GPA: 27.7/30
- Final mark: 109/110
- Thesis: "Particle Swarm Algorithm for Sphere Packing Problems"
- Adviser: Prof. Fabio Schoen

Employment

Microsoft Research – Advanced Technology Labs Israel Herzliya, Israel

Research intern Sep 2015 - Dec 2015

- Studied the problem of identifying small cohorts of search engine users who might be affected by the same disease.

MedStar Institute for Innovation (MI2) Washington, DC, USA

Summer intern May 2015 - Aug 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.

Teaching Experience

Georgetown University Washington, DC, USA

Teaching Assistantship (TA) August 2013 - Present

- Information Retrieval (undergraduate & graduate): Fall 2013, Fall 2014, Fall 2016.
- Introduction to Information Systems (undergraduate): Spring 2014.
- Introduction to Data Mining (undergraduate): Spring 2014, Spring 2015, Spring 2016, Spring 2017.
- Introduction to Database (undergraduate): Spring 2015.
- Health Search and Mining (graduate): Spring 2017.

Peer-Reviewed Publications

[8] L. Soldaini and E. Yom-Tov. "Inferring Individual Attributes from Search Engine Queries and Auxiliary Information" *Wide World Web conference (WWW)*. 2017.

[7] L. Soldaini and N. Goharian. "Learning to Rank for Consumer Health Search: a Semantic Approach." *European Conference on Information Retrieval (ECIR)*. 2017.

[6] L. Soldaini and N. Goharian. "QuickUMLS: a Fast, Unsupervised Approach for Medical Concept Extraction." *MedIR workshop*, *ACM conference on Research and Development in Information Retrieval (SIGIR)*. 2016.

[5] A. Cohan, L. Soldaini, and N. 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.

[4] L. Soldaini, A. Yates, E. Yom-Tov, O. Frieder, and N. Goharian. "Enhancing Web Search in the Medical Domain via Query Clarification." *Information Retrieval Journal*, April. 2016, volume 19, issue 1, Springer.

[3] A. Cohan, L. Soldaini, and N. 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.

[2] L. Soldaini, A. Cohan, A. Yates, N. Goharian, and O. Frieder. "Retrieving Medical Literature for Clinical Decision Support." *European Conference on Information Retrieval (ECIR)*. 2015.

[1] A. Cohan, L. Soldaini, A. Yates, N. Goharian, and O. Frieder. "On Clinical Decision Support." *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB)*. 2014.

Technical Reports

[3] L. Soldaini, W. Edman, N. 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)

[2] L. Soldaini, A. Cohan, A. Yates, N. Goharian, and O. Frieder. "Query Reformulation for Clinical Decision Support Search." *Text REtrieval Conference (TREC)*. 2014.

[1] A. Cohan, L. Soldaini, Saket S.R. Mengle, and N. Goharian. "Towards Citation-Based Summarization of Biomedical Literature." *Text Analysis Conference (TAC)*. 2014.

Professional Activities

- **Program Committee Member**, Computational Health, *WWW*. 2017.
- **Subreviewer**, *AAAI*. 2017.

Awards

- **Student Travel Grant**. *MedIR workshop. SIGIR* 2016.
- **Second Place at Best Poster Award** (2 out of 40). "On Clinical Decision Support". *ICBI Biomedical Informatics Symposium at Georgetown University* 2014.

Technical Skills

- **Programming languages**: Python (advanced), Bash (competent), Javascript, Java, C# (some exposure);
- **Frameworks**: web servers (Flask), databases (MySQL, MongoDB), virtualization (Docker, Vagrant), search engines (Elasticsearch, Terrier), machine learning (Keras, NumPy, scikit-learn, SciPy, Theano, Vopal Wabbit, Weka), natural language processing (spaCy, Stanford CoreNLP);
- **Platforms**: UNIX (OS X, RedHat Linux, Debian), Microsoft Windows.