Dana Movshovitz-Attias, PhD

CONTACT INFORMATION

Google Research
1600 Amphitheatre Pkwy
Mountain View, CA 94043 USA

E-mail: attias.dana@gmail.com
Website: www.cs.cmu.edu/~dmovshov

RESEARCH INTERESTS

I am interested in the intersection of Natural Language Processing, Information Retrieval, and Machine Learning. My research experience includes the following topics: knowledge base construction, grounded language learning, learning semantic relations, topic models, mining software repositories and software-focused corpora, bootstrapping on biomedical ontologies, knowledge base population, ontology learning, bootstrap learning and semantic drift, seed set refinement, text alignment with Hidden Markov Models, social media analysis, and computational biology.

EDUCATION

Carnegie Mellon University, Computer Science Department, Pittsburgh PA, USA

Ph.D., Computer Science, 08/2010-09/2015

- Thesis: *Grounded Knowledge Bases for Scientific Domains*
- Adviser: Professor William W. Cohen
- Research areas: Natural Language Processing, Information Retrieval, Machine Learning
- My research projects include:
  - Knowledge base and ontology construction for the software and biomedical domains.
  - Grounded learning of semantic relations.
  - Analysis of software source code using natural language models.
  - Detecting geo-location of entities using the NELL knowledge base.
- GPA: 4.0

Hebrew University of Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel

M.Sc., Computer Science and Computational Biology, 10/2007-07/2010

- *Magna Cum Laude*, in Computer Science and Computational Biology
- Thesis: *On the use of structural templates for high-resolution docking*
- Adviser: Professor Ora Schueler-Furman
- Research areas: Computational Biology, Protein-protein docking, Peptide-protein interactions
- Main project: High-resolution prediction of backbone flexibility in protein-protein docking complexes, using evolutionarily related proteins as structural templates. This was done within the framework of the RosettaDock software.
- Secondary project: Investigating the structural characteristics of peptide-protein interfaces using statistical models, and predicting peptide-specific binding sites. My models incorporate geometric, chemical and energetic properties into a unified classification framework.
- GPA: 94.5/100

B.Sc., Computer Science and Computational Biology, 10/2004-07/2007

- Minor in Math
- Adviser: Professor Ora Schueler-Furman
- Research areas: Computational Biology, Protein-protein docking
- Main project: High-resolution prediction of backbone flexibility in protein-protein docking complexes, using evolutionarily related proteins as structural templates. This was done within the framework of the RosettaDock software.
- GPA: 90.3/100

PROFESSIONAL EXPERIENCE

Google, Mountain View, California, USA — November 2015 - Present
Software Engineer at Google Research

Google, Mountain View, California, USA May 2014 - August 2014
Software Engineering Intern

- Hosts: Steven Whang and Alon Halevy, Structured Data Team at Google Research
- My project was focused on building a hierarchy between concepts, mainly based on concept attributes. For example, if the concepts 'university' and 'organization' both have the attribute 'members' this indicates a relation between them. Combining this data with other semantic and contextual information allows us to discover sub-concept to super-concept relations.
- This internship project was published as "Discovering Subsumption Relationships for Web-Based Ontologies" at WebDB 2015 and won a Best Paper Award.

Facebook, Menlo Park, California, USA May 2012 - August 2012
Software Engineering Intern

- Mentor: Eric Sun, Entities Team
- The Entities team is responsible for understanding entities and their relations in the Facebook graph. My project was focused on making inferences on relations between a person and their place of work, or where they attended school. This included evaluating existing relations and predicting new ones. In some cases, we identified conditions that lead to common errors in links, which we could automatically fix. The work involved analyzing large scale data using HIVE.

IBM Israel Software Labs (ILSL), Jerusalem, Israel August 2006 - June 2010
Software Engineer

- Developed algorithms and GUI for ICM (IBM Classification Module), a framework for automatic organization of unstructured natural language data, including document topic classification and clustering. I gained experience in multi-threaded programming, contributed to the core algorithms (in Java and C++), and wrote GUI in MFC and SWING.
- Developed dynamic web interfaces for ODE (OmniFind Discovery Edition), a JAVA-based search and navigation system, allowing end users to search using keywords or complex natural language queries.

Exent Technologies, Petach-Tikva, Israel October 2003 - October 2004
QA Engineer

- Exent provides application-on-demand services. The application is packaged in a format which allows delivering only needed portions to the user. My job was to simulate an end-user of the system, and find errors in the packaged software.

Military Service

Technology Center, Israeli Defense Force, Israel October 1999 - May 2003
Computer Programmer

- Software developer of remote management and control systems written in MFC, C++ and Java. My team developed control frameworks of both hardware and software, which could be easily expanded to managing additional applications, with minimal programming effort on the client side. Using these frameworks, we provided remote management capabilities to dozens of teams and systems throughout the organization.
- In the last year of my service, I lead the design and development of a new software management system. In that year, we completed the development of the system and setup service with initial users.

Publications

Movshovitz-Attias, D. and Cohen, W.W. (2015). KB-LDA: Jointly Learning a Knowledge Base of Hierarchy, Relations, and Facts. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*. Association for Computational Linguistics.

Movshovitz-Attias, D., Whang, S.E., Noy, N., and Halevy, A. (2015). Discovering Subsumption Relationships for Web-Based Ontologies. In *Proceedings of the 18th*

International Workshop on the Web and Databases (WebDB). ACM. Winner of the Best Paper Award at WebDB.

Movshovitz-Attias, D. and Cohen, W.W. (2015). Grounded Discovery of Coordinate Term Relationships between Software Entities. *arXiv preprint arXiv:1505.00277*.

Movshovitz-Attias, D. and Cohen, W.W. (2013). Natural Language Models for Predicting Programming Comments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. ACL, Sofia, Bulgaria.

Movshovitz-Attias, D., Movshovitz-Attias, Y., Steenkiste, P., and Faloutsos, C. (2013). Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow. In *ASONAM*. Niagara Falls, Canada.

Lavi, A., Ngan, C.H., Movshovitz-Attias, D., Bohnuud, T., Yueh, C., Beglov, D., Schueler-Furman, O., and Kozakov, D. (2013). Detection of peptide-binding sites on protein surfaces: The first step towards the modeling and targeting of peptide-mediated interactions. *Proteins: Structure, Function, and Bioinformatics*.

Movshovitz-Attias, D. and W. Cohen, W. (2012). Bootstrapping Biomedical Ontologies for Scientific Text using NELL. In *BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, in NAACL*, pp. 11–19. Association for Computational Linguistics, Montréal, Canada.

Movshovitz-Attias, D. and W. Cohen, W. (2012). Alignment-HMM-based Extraction of Abbreviations from Biomedical Text. In *BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, in NAACL*, pp. 47–55. Association for Computational Linguistics, Montréal, Canada.

Movshovitz-Attias, D. and W. Cohen, W. (2012). Bootstrapping Biomedical Ontologies for Scientific Text using NELL. *Carnegie Mellon University MLD Technical Report: CMU-ML-12-101*.

London, N., Raveh, B., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). Can Self-Inhibitory Peptides be Derived from the Interfaces of Globular Protein–Protein Interactions? *Proteins: Structure, Function, and Bioinformatics*.

Movshovitz-Attias, D., London, N., and Schueler-Furman, O. (2010). On the use of structural templates for high-resolution docking. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1939–1949.

London, N., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. *Structure*, 18(2):188–199.

Talks

Understanding Software Language November 2015
Seminar on Programming with "Big Code", Schloss Dagstuhl, Germany
Talk presented by William Cohen

Knowledge Base Construction for the Software Domain September 2014
University of California, Davis

Analyzing Software Source Code using Natural Language Models May 2013
University of California, Davis

Never Ending Learning from Biomedial Text January 2012
Google Grad CS Forum, San Francisco, California

Semi-Supervised Learning in the Biomedical Domain April 2011
Worldly Knowledge collaboration meeting with Google Pittsburgh, at CMU

Teaching Experience

Language and Statistics (11-761), Carnegie Mellon University — Spring 2014
Teaching Assistant for Prof. Roni Rosenfeld

Artificial Intelligence: Representation and Problem Solving (15-381), Carnegie Mellon University — Fall 2012
Teaching Assistant for Prof. Ariel Procaccia and Prof. Emma Brunskill

Professional Service

Workshop Organizing Committee

- NSF Interdisciplinary Workshop on Statistical NLP and Software Engineering, hosted at Microsoft Research, Redmond, WA — October 2015

Conference Reviewing

- Association for Computational Linguistics (ACL) — 2014
- Empirical Methods in Natural Language Processing (EMNLP) — 2014, 2012

Doctoral Review Committee

- Computer Science Department at Carnegie Mellon University — 2015, 2014, 2013

Posters

BioNELL: Bootstrapping Biomedical Ontologies for Scientific Text using NELL

- Knowledge Discovery Workshop, Google NYC — 2011
- Worldly Knowledge collaboration meeting with Google Pittsburgh — September 2011

The structural basis of peptide-protein binding strategies

- 12th Israeli Bioinformatics Symposium at Weizmann Institute, Israel — 2009

On the use of structural templates for high-resolution docking

- 11th Israeli Bioinformatics Symposium at Tel-Aviv University, Israel — 2008

Competitions

CAPRI: Critical Assessment of PRediction of Interactions

- Participant in rounds 15-17 of the CAPRI experiment for computational prediction of protein complex structures. I was part of the joint RosettaDock team of Prof. David Baker from the University of Washington, and Dr. Ora Schueler-Furman from the Hebrew University.

Awards

- Best Paper Award at WebDB for "Discovering Subsumption Relationships for Web-Based Ontologies". 2015

Hebrew University of Jerusalem

- Computer Science and Computational Biology program, Tuition Scholarship for Graduate Achievements. 2009
- Rector's Scholarship for Undergraduate Achievements. 2008
- Dean's Honors List. 2006, 2007
- Computer Science and Computational Biology, Stein Scholarship. 2005, 2006

Volunteer Work

Google I/O Youth, Google I/O 2014

- Volunteered at Google I/O for the I/O Youth program, introducing high-school children, ages 11-15, to Google I/O. This is a day of interactive, project-based sessions to introduce the children to practical applications for coding, robotics, and computer music. I mentored coding sessions in CoffeeScript (based on JQuery), guiding children in their projects.

Woman@SCS, Carnegie Mellon University

- Participated in the Big Sister program as a mentor to CS undergraduate students (2010, 2013, 2014).
- Represented CMU at the AAAS meeting of science and technology, at Washington D.C., introducing participating children to Computer Science with problem solving activities (2010).
- Performed in the Women@SCS Outreach Roadshow. The show introduces middle school children to basic Computer Science concepts by an interactive performance (2010).

LANGUAGES

- Hebrew: Native language.
- English: Fluent.
- French: Basic competence.