

Ariel S. Schwartz

Department of Bioengineering
University of California, San Diego
PFBH Rm 202-4
9500 Gilman Drive, Mail Code 0412
La Jolla, CA 92093-0412
USA

Phone: +1 (858) 822-4705
sariel@ucsd.edu
<http://chianti.ucsd.edu/~sariel>

Education

Ph.D. Computer Science, March 2007
University of California, Berkeley
Advisors: Lior Pachter and Marti Hearst

B.Sc. Information Systems Engineering, December 1999
Computer Science Department, Technion—Israel Institute of Technology
Summa Cum Laude
GPA: 95/100

Research Interests

Comparative genomics, multiple sequence alignment algorithms, systems biology, graphical models, posterior decoding algorithms, text mining.

Research Experience

2007–Postdoctoral Scholar
Trey IdekerUniversity of California, San Diego
Design and analysis of experimental techniques towards improved coverage and accuracy of protein-protein interaction networks. Analysis of combinatorial regulation in tissue-specific transcription factor networks.

2003–2006Research Assistant
Lior Pachter and Eugene W. MyersUniversity of California, Berkeley
Developed novel algorithms for multiple sequence alignment and comparative gene finding. Defined a metric for the space of sequence alignments, and a new alignment accuracy measure based on this metric. Designed and implemented *AMAP*, a multiple sequence alignment program, based on posterior-decoding, and a novel *sequence annealing* algorithm for improved alignment accuracy.

2002–2006Research Assistant
Marti HearstUniversity of California, Berkeley
Designed and implemented the *BioText* system's architecture. Lead the implementation of *LQL*, a new language and infrastructure for search over large collections of syntactic and semantic text annotations. Participated in the *TREC* genomics track. Designed text-mining algorithms for the biomedical domain, including an abbreviation recognition algorithm, and a posterior decoding algorithm for alignment of citation sentences.

Publications

A. S. Schwartz, Posterior Decoding Methods for Optimization and Accuracy Control of Multiple Alignments, PhD dissertation, *EECS Department, University of California, Berkeley*, Technical Report No. UCB/EECS-2007-39, March 28, 2007.

A. S. Schwartz, A. Divoli and M. Hearst, Multiple Alignment of Citation Sentences with Conditional Random Fields and Posterior Decoding, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Prague, Czech Republic, June 2007, p 847–857.

A. S. Schwartz and L. Pachter, Multiple Alignment by Sequence Annealing, *Bioinformatics* 23 (2007), e24–e29.

A. S. Schwartz, E. W. Myers and L. Pachter, Alignment Metric Accuracy, *arXiv:q-bio/0510052*.

E.H. Margulies et al., Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome , *Genome Research*, 2007 17: 760–774.

P. Nakov, A. Schwartz, B. Wolf and M. Hearst, Scaling Up BioNLP: Application of a Text Annotation Architecture to Noun Compound Bracketing, in *ACL/ISMB BioLINK SIG: Linking Literature, Information and Knowledge for Biology*, Detroit, MI, June 2005.

P. Nakov, A. Schwartz, B. Wolf and M. Hearst, Supporting Annotation Layers for Natural Language Processing, in *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, Ann Arbor, MI, June 2005, p 65–68.

P. I. Nakov, A. S. Schwartz, E. Stoica and M. A. Hearst, BioText Team Experiments for the TREC 2004 Genomics Track, in *Proceedings of TREC 2004* Gaithersburg, MD, 2005.

P. I. Nakov, A. Schwartz and M. A. Hearst, Citances: Citation Sentences for Semantic Analysis of Bioscience Text, in *the SIGIR'04 Workshop on Search and Discovery in Bioinformatics*, Sheffield, UK, July 2004.

D. E. Oliver, G. Bhalotia, A. S. Schwartz, R. B. Altman and M. A. Hearst, Tools for Loading MEDLINE into a Local Relational Database, *BMC Bioinformatics*, 5 (146), Oct 2004.

G. Bhalotia, P. I. Nakov, A. S. Schwartz and M. A. Hearst, BioText Team Report for the TREC 2003 Genomics Track, in *Proceedings of TREC 2003*, Gaithersburg, MD, 2004.

A. S. Schwartz and M. A. Hearst, A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text, in *the Proceedings of the Pacific Symposium on Biocomputing*, 8:451-462(2003).

Invited Talks

Posterior Decoding Methods for Optimization and Accuracy Control of Multiple Alignments, *UCSD—CSE—Bioinformatics Seminar*, San Diego, CA, April 2007.

Posterior Decoding Methods for Comparative Sequence Analysis, *MSRI Summer Graduate Workshop on Mathematical aspects of computational biology*, Berkeley, CA, June 2006.

Multiple Alignment by Sequence Annealing, *Second Annual Retreat of the DE Graduate Group in Computational and Genomic Biology*, Bodoga Bay, CA, April 2006.

AMAP - Posterior Decoding based Algorithm for Improved Sequence Alignment Accuracy and Sensitivity/Specificity Tradeoff Control, *CIPRES All Hands Meeting*, Austin, TX, February 2006.

Demonstrations

AMAP—Fast and accurate multiple alignment using posterior decoding and sequence annealing, *15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) & 6th European Conference on Computational Biology (ECCB)*, Vienna, Austria, July, 2007.

Supporting Annotation Layers for Natural Language Processing, *4^{3^d} Annual Meeting of the Association for Computational Linguistics (ACL'05)*, Ann Arbor, MI, June 2005.

Data and Software

AMAP - fast and accurate amino acid multiple sequence alignment program, which utilizes posterior decoding, and a novel sequence annealing alignment method. It is the only alignment program that allows to control the sensitivity / specificity tradeoff. In its default configuration, *AMAP* is tuned to maximize the expected *Alignment Metric Accuracy (AMA)* score - a new alignment accuracy measure, based on a metric for the multiple-alignment space, which integrates sensitivity and specificity into a single balanced measure.

ENCODE Alignment agreement custom tracks for the UCSC Genome Browser.

Layered Query Language (LQL) - Infrastructure and query language for layers of annotated text.

Abbreviation recognition software.

Software to parse and load MEDLINE into an RDBMS.

Awards and Honors

Ruth L. Kirschstein National Research Service Award (NRSA), UCSD Bioengineering NIH training grant, 2007-2008.

Best paper award, ECCB 2006.

GAANN Fellowship recipient, University of California, Berkeley, Computer Science Division, 8/2001-7/2002, 1/2003-7/2003.

Graduated *Summa Cum Laude*, Technion - Israel Institute of Technology, Computer Science Department, 1999.

Provost's List, Technion - Israel Institute of technology, 1996-1999.

Outstanding Student Award, Technion - Israel Institute of technology, Industrial Engineering and Management Department, 1998.

Teaching Experience

Fall 2002 Teaching Assistant
Introduction to Database Systems University of California, Berkeley
Prepared and taught discussion sections in an undergraduate database class, created and managed course projects, and helped prepare and grade exams.

Fall 1998 Teaching Assistant
Deterministic Models in Technion - Israel Institute of Technology
Operations Research
Taught discussion sections in an undergraduate operation research class, graded course projects.

Employment

3-6/2002 Information Systems Consultant
Haas School of Management University of California, Berkeley
Analyzed the school's alumni donation tracking systems, and its interface to the university's system.

1/1998-7/2001 Information Systems Engineer
Information Systems Group Intel Israel Development Center
Designed and implemented decision support systems, and project management systems, including leading the design and implementation of a distributed data warehouse project for managing the IT resources of Intel's development centers around the world.

3/1992-3/1996 Tank Commander and Intelligence Officer
Israel Defense Forces Israel
Classified.

Graduate Coursework

Computer Science

Advanced topics in database systems, knowledge representation and reasoning, natural language processing, statistical learning theory, combinatorial algorithms and data structures, advanced topics in learning and decision making.

Computational Biology

Protein informatics, algorithms for computational biology, statistical genetics.

Biological Sciences

General biochemistry and molecular biology, general genetics, advanced genetics.

References

Professor Lior Pachter
1081 Evans Hall
Department of Mathematics
University of California, Berkeley
Berkeley, CA 94720-3840
Phone: +1 (510) 642-2028
Fax: +1 (510) 642-8204
lpachter@math.berkeley.edu

Professor Marti Hearst
102 South Hall
School of Information
University of California, Berkeley
Berkeley, CA 94720-4600
Phone: +1 (510) 642-8016
Fax: +1 (510) 642-5814
hearst@sims.berkeley.edu

Professor Trey Ideker
9500 Gilman Drive, Mail Code 0412
Department of Bioengineering
University of California, San Diego
La Jolla, CA 92093-0412
Phone: +1 (858) 822-4558
Fax: +1 (858) 822-4246
trey@bioeng.ucsd.edu

Professor Eugene W. Myers
Janelia Farm Research Campus
Howard Hughes Medical Institute
19700 Helix Drive
Ashburn, VA, 20147-2408
Phone: +1 (571) 209-4153
Fax: +1 (571) 209-4083
myersg@janelia.hhmi.org